Biometrics 000, 1–21 DOI: 000
000 0000

# Quantile Regression for Recurrent Gap Time Data

**Xianghua Luo[1,*], Chiung-Yu Huang[2], and Lan Wang[3]**

[1] Division of Biostatistics, School of Public Health,
University of Minnesota, Minneapolis, MN 55455, U.S.A.

[2] Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases,
National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

[3] School of Statistics, University of Minnesota, Minneapolis, MN 55455, U.S.A.

*email:* luox0054@umn.edu

Summary: Evaluating covariate effects on gap times between successive recurrent events is of interest in many medical and public health studies. While most existing methods for recurrent gap time analysis focus on modeling the hazard function of gap times, a direct interpretation of the covariate effects on the gap times is not available through these methods. In this paper, we consider quantile regression that can provide direct assessment of covariate effects on the quantiles of the gap time distribution. Following the spirit of the weighted risk-set method by Luo and Huang (2011, *Statistics in Medicine* **30,** 301–311), we extend the martingale-based estimating equation method proposed by Peng and Huang (2008, *Journal of American Statistical Association* **103,** 637–649) for univariate survival data to recurrent gap time data. The proposed estimation procedure can be easily implemented in existing software packages for univariate censored quantile regression. Uniform consistency and weak convergence of the proposed estimators are established. Monte-Carlo studies demonstrate the effectiveness of the proposed method. An application to data from the Danish Psychiatric Central Register is presented to illustrate the theory and methods developed in this paper.

Key words: Clustered survival data; Data perturbation; Gap times; Quantile regression; Recurrent events; Within-cluster resampling.

## 1. Introduction

Recurrent event data frequently arise in medical and epidemiology studies, where each subject may experience a number of "failures" over the course of follow-up. Examples of recurrent failure events include tumor recurrences, recurrent infections, and repeated hospitalization. The times between successive recurrent events (gap times) are the natural outcome of interest in these studies. Most existing methods for recurrent gap time analysis focus on modeling the hazard functions of gap times; for example, Huang and Chen (2003) studied proportional hazards (PH) models and Sun, Park, and Sun (2006) considered additive hazards models. Other researchers have considered linear models for the covariate effects on the (transformed) gap times. For example, Chang (2004) and Strawderman (2005) proposed accelerated failure time (AFT) models for logarithm transformed gap times, and Lu (2005) proposed semiparametric linear transformation models by assuming linear covariate effects on an unspecified monotone increasing function of the gap times. However, these models only consider the effects of the covariates on the mean of the (transformed) gap time, thus important forms of heterogeneity might be overlooked. Furthermore, the associated theory generally requires that the data are homoscedastic in the sense that the error terms in the regression models are independent of the covariates. This assumption is often found to be too restrictive in practice.

In this paper, we consider quantile regression for recurrent gap time data. Formulation of quantile regression is based on the conditional quantiles of the gap times given the covariates of interest. The covariate effects in quantile regression models are easier to interpret than those of hazard-based models. In addition, contrary to the PH model, quantile regression does not need the PH assumption and hence is more flexible. More importantly, quantile regression allows the covariates to have different effects at the tails or at different segments of the conditional distribution of the gap times and therefore allows for heteroscedasticity

in the data. Quantile regression provides a complete picture on the covariate effects on the gap time. This modeling framework is also naturally robust for heavy-tailed gap times.

Quantile regression was originally proposed by Koenker and Bassett (1978) for analyzing uncensored continuous data. For univariate survival data, quantile regression was first studied by Powell (1984, 1986) for data with fixed censoring where the censoring times are observed for all observations. The random censoring case was first studied by Ying, Jung, and Wei (1995) which assumed that the event time and the censoring time are unconditionally independent, and by Yang (1999) under a more relaxed conditional independence assumption but with independent and identically distributed (i.i.d.) error distributions. Portnoy (2003) proposed a novel recursively reweighted estimator of the censored regression quantile process, which is a generalization of the Kaplan-Meier estimator. Recently, Peng and Huang (2008) proposed a martingale-based estimating equation approach that can be viewed as an extension of Nelson-Aalen type estimation and displays similar performance as Portnoy's estimator; Wang and Wang (2009) proposed a locally weighted estimating procedure that relaxes the global linearity assumption.

In this paper, we extend the martingale-based estimating equation approach by Peng and Huang (2008) for univariate survival data to recurrent gap time data following the spirit of the weighted risk-set method proposed by Luo and Huang (2011). The proposed estimation procedure can be easily implemented in existing software packages for univariate censored quantile regression. Applying the theory of empirical processes, we establish uniform consistency and weak convergence for the proposed estimator of the regression quantiles.

The rest of the article is organized as follows. In Section 2 we first present the estimation procedure for the first gap time only data by using Peng and Huang's method. We then discuss how this method can be extended to recurrent gap time data. We propose to use the bootstrap resampling and data perturbation methods for estimating the variance of

the proposed estimator. In Section 3 we report the results from simulation studies for both homoscedastic and heteroscedastic recurrent event data. In Section 4 we analyze the hospitalization data from the Danish Psychiatric Central Register. Some concluding remarks are provided in Section 5.

## 2. Method

### 2.1 *Model Setup*

We begin by introducing some notation to describe the structure of the recurrent gap time data. Suppose $n$ subjects are recruited into a study after experiencing an initial event. Let $i = 1, 2, \ldots, n$ index subjects and $j = 0, 1, 2, \ldots$ index the sequence of the recurrent events for a subject, where $j = 0$ indexes the initial event. For the $i$th subject, $T_{ij}$ denotes the gap time between the $(j-1)$th and $j$th events and $C_i$ denotes the time between enrollment and the end of follow-up. Let $m_i$ be the index of the last censored gap time, so that $m_i$ satisfies the constraint $\sum_{j=1}^{m_i-1} T_{ij} \leqslant C_i$ and $\sum_{j=1}^{m_i} T_{ij} > C_i$, where $\sum_1^0 \cdot \equiv 0$. Define $\Delta_i = I(m_i > 1)$, where $I(\cdot)$ is the indicator function. Thus $\Delta_i = 0$ if the subject is free of events during follow-up, and $\Delta_i = 1$ if the subject experienced recurrent events. As discussed in Wang and Chang (1999), the unique structure of recurrent events generates many difficulties in modeling gap time data. Firstly, the second and later gap times are subject to "induced" dependent censoring. Specifically, while the first gap time, $T_{i1}$ may be subject to independent censoring $C_i$, the second and later gap times, $T_{ij} (j \geqslant 2)$, are subject to dependent censoring $C_i - T_{i1} - \ldots - T_{i,j-1}$, as gap times of the same subject are usually correlated. Secondly, due to intercept sampling, the last censored gap time, $T_{im_i}$, tends to be longer than the uncensored gap times. Finally, the number of gap times is informative about the underlying distribution, as subjects at a higher risk of experiencing recurrent events are likely to have

shorter, and hence more, gap times. Because of these reasons, naively treating recurrent gap time data as clustered survival data (e.g. family data) would not be appropriate.

[Figure 1 about here.]

Let $\boldsymbol{Z}_i$ be a $p \times 1$ vector of time-independent covariates, including the intercept. The $\tau$th conditional quantile of $T_{ij}$ given $\boldsymbol{Z}_i$ is defined as $Q_{ij}(\tau|\boldsymbol{Z}_i) = \inf_t\{\Pr(T_{ij} \leqslant t|\boldsymbol{Z}_i) \geqslant \tau\}$, for $\tau \in (0,1)$. To model the effect of $\boldsymbol{Z}_i$ on the quantiles of $T_{ij}$, we impose the following model assumptions:

M1: There exists a nonnegative subject-specific frailty variable (or vector) $\gamma_i$ such that, conditioning on $\gamma_i$ and $\boldsymbol{Z}_i$, the individual recurrent event process is a renewal process.

M2: Given $\boldsymbol{Z}_i$, the conditional quantiles of $T_{ij}$ satisfy the quantile regression model

$$Q_{ij}(\tau|\boldsymbol{Z}_i) = \exp\{\boldsymbol{Z}_i^{\intercal}\boldsymbol{\beta}(\tau)\} \text{ for } \tau \in (0,1), \tag{1}$$

where $\boldsymbol{\beta}(\tau)$ is a $p$-dimensional vector of quantile coefficients and may change with $\tau$.

M3: Given $\boldsymbol{Z}_i$, the censoring time $C_i$ is independent of $\gamma_i$ and $(T_{i1}, T_{i2}, \ldots)$.

Assumption M1 implies that, conditioning on $\gamma_i$ and $\boldsymbol{Z}_i$, the underlying recurrent gap times $T_{i1}, T_{i2}, \ldots$ are i.i.d. As a result, by integrating out $\gamma_i$, the gap times $T_{i1}, T_{i2}, \ldots$ are identically (but not independently) distributed. The conditional i.i.d. assumption is a weaker condition than the i.i.d. assumption imposed by other competing appraoches, including Strawderman (2005). We note that both the distribution of the subject-specific frailty $\gamma_i$ and the dependency between $\gamma_i$ and $(T_{i1}, T_{i2}, \ldots)$ are left unspecified under the model assumptions. This is contrast to the conventional shared-frailty models for correlated survival data, such as Huang and Liu (2007), where a parametric assumption for the frailty distribution is usually required for statistical inference.

We further note that Assumption M2 specifies the marginal distribution of the gap times unconditional on $\gamma_i$, hence the proposed model in Equation (1) can be considered as a

marginal model. It is easy to see that the proposed model reduces to the AFT model for recurrent gap times studied by Chang (2004) when the regression quantile coefficients corresponding to the non-intercept covariates are invariant with $\tau$.

Define the observed gap time $X_{ij} = T_{ij}$ for $j < m_i$ and $X_{im_i} = C_i - \sum_{j=1}^{m_i-1} T_{ij}$. Denote the observed data for subject $i$ by $\{\boldsymbol{Z}_i, \Delta_i, X_{i1}, \ldots, X_{im_i}\}$. The observed data from the $n$ subjects are assumed to be i.i.d. For convenience, we define $m_i^* = \max\{m_i - 1, 1\}$. Hence, for people with no observed events ($\Delta_i = 0$), $m_i^* = 1$; and for people with observed events ($\Delta_i = 1$), $m_i^*$ is the number of uncensored gap times (see Figure 1).

### 2.2 *Weighted Risk-Set Estimators*

To estimate $\boldsymbol{\beta}(\tau)$, a simple approach would be applying the estimation procedure for quantile regression for univariate survival data to the subset of the data that comprised of only the first gap times, $\{\boldsymbol{Z}_i, \Delta_i, X_{i1}\}, i = 1, 2, \ldots, n$. In particular, we apply the martingale-based estimation equation proposed by Peng and Huang (2008) and estimate the quantile regression model by solving $\boldsymbol{U}(\boldsymbol{\beta}, \tau) = \boldsymbol{0}$, where

$$\boldsymbol{U}(\boldsymbol{\beta}, \tau) = n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i \left( N_{i1} \left[ \exp\{\boldsymbol{Z}_i^{\intercal} \boldsymbol{\beta}(\tau)\} \right] - \int_0^{\tau} R_{i1} \left[ \exp\{\boldsymbol{Z}_i^{\intercal} \boldsymbol{\beta}(u)\} \right] dH(u) \right), \tag{2}$$

with $N_{ij}(t) = I(X_{ij} \leqslant t, \Delta_i = 1)$, $R_{ij}(t) = I(X_{ij} \geqslant t)$, and $H(x) = -\log(1-x)$ for $0 \leqslant x < 1$. The time to first event analysis, however, is expected to be inefficient because the second and later gap times are not used in the formulation of (2).

For recurrent gap time data, Luo and Huang (2011) introduced two weighted stochastic processes as important building blocks for estimation procedure, namely, the averaged counting process $N_i^*(t) = (m_i^*)^{-1} \sum_{j=1}^{m_i^*} N_{ij}(t)$ and the averaged at-risk process $R_i^*(t) = (m_i^*)^{-1} \sum_{j=1}^{m_i^*} R_{ij}(t)$. Note that $N_i^*(t)$ and $R_i^*(t)$ have a jump size $1/m_i^*$ and $-1/m_i^*$, respectively, at each uncensored gap time. This is different from the univariate survival analysis where the jump size of the counting processes $N_{i1}(t)$ and the at-risk processes $R_{i1}(t)$ is always 1 and $-1$. In the same spirit as the weighted risk-set (WRS) method for the PH

regression discussed in Luo and Huang (2011), we replace $N_{i1}$ and $R_{i1}$ in $\boldsymbol{U}(\boldsymbol{\beta}, \tau)$ with their counterparts $N_i^*$ and $R_i^*$ to formulate a new estimating equation

$$\boldsymbol{U}^*(\boldsymbol{\beta}, \tau) = n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i \left[ N_i^* \left( \exp\{\boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(\tau)\} \right) - \int_0^\tau R_i^* \left( \exp\{\boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(u)\} \right) dH(u) \right], \tag{3}$$

and estimate $\boldsymbol{\beta}(\tau)$ by solving the WRS estimating equation $\boldsymbol{U}^*(\boldsymbol{\beta}, \tau) = \mathbf{0}$. Note that under Assumptions M1 and M3, the observed uncensored gap times $X_{i1}, \ldots, X_{im_i^*}$ are identically distributed conditional on $(\gamma_i, \boldsymbol{Z}_i, m_i, C_i)$. Hence, by double expectation we can show that $\mathrm{E}[N_i^*(t) \mid \boldsymbol{Z}_i] = \mathrm{E}[N_{i1}(t) \mid \boldsymbol{Z}_i]$ and $\mathrm{E}[R_i^*(t) \mid \boldsymbol{Z}_i] = \mathrm{E}[R_{i1}(t) \mid \boldsymbol{Z}_i]$. As a result, we have $\mathrm{E}[\boldsymbol{U}^*(\boldsymbol{\beta}, \tau)] = \mathrm{E}[\boldsymbol{U}(\boldsymbol{\beta}, \tau)]$, which ensures the consistency of the proposed estimator.

Next, we show how the proposed method can be implemented in existing software packages for univariate quantile regression. Note that the WRS estimating equation $\boldsymbol{U}^*(\boldsymbol{\beta}, \tau)$ can be reexpressed as

$$\boldsymbol{U}^*(\boldsymbol{\beta}, \tau) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i^*} \frac{\boldsymbol{Z}_i}{m_i^*} \left[ \Delta_i I \left\{ \frac{\log X_{ij}}{m_i^*} \leqslant \frac{\boldsymbol{Z}_i^\intercal}{m_i^*} \boldsymbol{\beta}(\tau) \right\} - \int_0^\tau I \left\{ \frac{\log X_{ij}}{m_i^*} \geqslant \frac{\boldsymbol{Z}_i^\intercal}{m_i^*} \boldsymbol{\beta}(\tau) \right\} dH(u) \right].$$

This suggests that we can assign a weight $w_{ij} = 1/m_i^*$ to the first $m_i^*$ observations of subject $i$ to obtain working data $\{(\Delta_i, \log(X_{ij}^{(w)}), \boldsymbol{Z}_i^{(w)}), i = 1, \ldots, n, j = 1 \ldots, m_i^*\}$, where

$$\log(X_{ij}^{(w)}) = w_{ij} \log(X_{ij}) \quad \text{and} \quad \boldsymbol{Z}_i^{(w)} = w_{ij} \boldsymbol{Z}_i.$$

Note that the censoring indicator $\Delta_i$ remains unchanged for each subject and the last censored gap time is discarded for those subjects who have at least one complete gap time after the data manipulation. We apply the martingale-based estimating equation method to the working data as if they were i.i.d. observations with sample size $\sum_{i=1}^n m_i^*$. This yields an estimating function

$$n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i^*} \boldsymbol{Z}_i^{(w)} \left( N_{ij}^{(w)} \left[ \exp\{\boldsymbol{Z}_i^{(w)\intercal} \boldsymbol{\beta}(\tau)\} \right] - \int_0^\tau R_{ij}^{(w)} \left[ \exp\{\boldsymbol{Z}_i^{(w)\intercal} \boldsymbol{\beta}(u)\} \right] dH(u) \right), \tag{4}$$

where $N_{ij}^{(w)}(t) = I(X_{ij}^{(w)} \leqslant t, \Delta_i = 1)$ and $R_{ij}^{(w)}(t) = I(X_{ij}^{(w)} \geqslant t)$. By some simple algebra, we have $N_{ij}^{(w)}\{\exp(\boldsymbol{Z}_i^{(w)\intercal} \boldsymbol{b})\} = N_{ij}\{\exp(\boldsymbol{Z}_i^\intercal \boldsymbol{b})\}$ and $R_{ij}^{(w)}\{\exp(\boldsymbol{Z}_i^{(w)\intercal} \boldsymbol{b})\} = R_{ij}\{\exp(\boldsymbol{Z}_i^\intercal \boldsymbol{b})\}$, and hence the working estimating equation (4) is equivalent to $\boldsymbol{U}^*$ in (3).

As in Peng and Huang (2008), we can use a grid-based estimation procedure to obtain an estimate of the true quantile coefficient $\boldsymbol{\beta}_0(\cdot)$. Specifically, we define an estimator of $\boldsymbol{\beta}_0(\tau)$, denoted by $\widehat{\boldsymbol{\beta}}^*(\tau)$, as a right-continuous piecewise-constant function that jumps only at $\tau_k, k = 1, \ldots, L$, where $0 < \tau_1 < \cdots < \tau_L = \tau_U < 1$ and $\tau_U$ is a constant subject to certain identifiability constraints due to censoring. We can obtain $\widehat{\boldsymbol{\beta}}^*(\tau_k)$ by sequentially solving the following estimating equation, which approximates (3), for $\boldsymbol{\beta}(\tau_k)$:

$$n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i^*}\boldsymbol{Z}_i^{(w)}\left[N_{ij}^{(w)}\left(\exp\left\{\boldsymbol{Z}_i^{(w)\intercal}\boldsymbol{\beta}(\tau_k)\right\}\right)-\sum_{l=0}^{k-1}R_{ij}^{(w)}\left(\exp\left\{\boldsymbol{Z}_i^{(w)\intercal}\widehat{\boldsymbol{\beta}}^*(\tau_l)\right\}\right)\right.$$
$$\left.\times\{H(\tau_{l+1})-H(\tau_l)\}\right]=0, k=1,\ldots,L, \qquad (5)$$

where $\exp\{\boldsymbol{Z}_i^{(w)\intercal}\widehat{\boldsymbol{\beta}}^*(0)\}$ for $i = 1, \ldots, n$ are set to be 0. Since equation (5) is not continuous, an exact root may not exist. Following Peng and Huang (2008), we define $\widehat{\boldsymbol{\beta}}^*(\tau_k)$ as the generalized solutions of (5), for which a slight perturbation of any of its components results in changing sign of the estimating function (Fygenson and Ritov, 1994). In fact, solving the estimating equation (5) for 0 is equivalent to locating the minimizer of the following $L_1$-type convex objective function:

$$\begin{aligned}\boldsymbol{V}_k(\boldsymbol{b}) &= \sum_{i=1}^{n}\sum_{j=1}^{m_i^*}\left|\Delta_i\log(X_{ij}^{(w)})-\Delta_i\boldsymbol{b}^\intercal\boldsymbol{Z}_i^{(w)}\right|+\left|A-\boldsymbol{b}^\intercal\sum_{i=1}^{n}\sum_{j=1}^{m_i^*}(-\Delta_i\boldsymbol{Z}_i^{(w)})\right|\\&+\left|A-\boldsymbol{b}^\intercal\sum_{i=1}^{n}\sum_{j=1}^{m_i^*}\left(2\boldsymbol{Z}_i^{(w)}\sum_{l=0}^{k-1}R_{ij}^{(w)}\left[\exp\{\boldsymbol{Z}_i^{(w)\intercal}\widehat{\boldsymbol{\beta}}^*(\tau_l)\}\right]\times\{H(\tau_{l+1})-H(\tau_l)\}\right)\right|, \qquad (6)\end{aligned}$$

where $A$ is a very large number. Alternatively, as argued in Peng and Huang (2008) and Koenker (2008a), finding the solution to (5) can be formulated as a linear programming problem. Therefore the estimation of $\boldsymbol{\beta}_0(\tau)$ also can be obtained by using software for linear programming problems, such as the R package, *quantreg* by Koenker (2008b).

We now establish the uniform consistency and weak convergence of the proposed estimator $\widehat{\boldsymbol{\beta}}^*(\tau)$. Let $\mathcal{S}_L = \{0 = \tau_0 < \tau_1 < \ldots < \tau_L = \tau_U < 1\}$ denote the grid in $\tau$, and let $\| \mathcal{S}_L \| = \max\{\tau_k - \tau_{k-1}; k = 1, \ldots, L\}$ denote the size of $\mathcal{S}_L$, where $\mathcal{S}_L$ may depend on $n$. For

a column vector $\boldsymbol{a}$, let $\|\boldsymbol{a}\|$ denote the Euclidean norm of $\boldsymbol{a}$. Under the regularity conditions (C1-C4) provided in Web Appendix A, we have the following theorem.

**Theorem 1.** Assuming that conditions M1-M3 and C1-C4 hold, provided $\lim_{n\to\infty} n^{1/2} \parallel \mathcal{S}_L \parallel = 0$, then $\sup_{\tau\in[\nu,\tau_U]} \parallel \widehat{\boldsymbol{\beta}}^*(\tau) - \boldsymbol{\beta}_0(\tau) \parallel \to 0$ in probability and $n^{1/2}\{\widehat{\boldsymbol{\beta}}^*(\tau) - \boldsymbol{\beta}_0(\tau)\}$ converges weakly to a Gaussian process for $\tau \in [\nu, \tau_U]$, where $0 < \nu < \tau_U$.

The proof of asymptotic properties closely follows the proof for the univariate survival data given in Peng and Huang (2008) and is provided in Web Appendices A and B. Note that identifiability is an inherent issue in the censored quantile regression. In principle, $\tau_U$ can be chosen according to the range of interest. In practice, $\tau_U$ may be obtained in an adaptive manner as the estimating equations in (5) are being solved sequentially. For example, when the minimization of the $k$th objective function $\boldsymbol{V}_k(\boldsymbol{b})$ is not feasible, it may suggest that the identifiability condition is not satisfied for all $\tau \geqslant \tau_k$. Hence, $\tau_U$ may be set to some value below the first $\tau_k$ in the sequence, $\{0 < \tau_1 < \tau_2 < \ldots\}$, for which the estimation of $\boldsymbol{\beta}(\tau_k)$ is not successfully carried out.

### 2.3 *Resampling Methods for Variance Estimation*

Both data perturbation and bootstrap resampling methods for the variance estimation of the regression quantiles are available in the R package *quantreg* for univariate censored quantile regressions. We propose to extend both methods to the variance estimation of the proposed estimator $\widehat{\boldsymbol{\beta}}^*(\tau)$ for recurrent gap time data. To apply the data perturbation method proposed by Jin et al. (2001) to our recurrent gap time data, we first generate a simple random sample of size $n$, $(\zeta_1, \ldots, \zeta_n)$ from a nonnegative distribution with unit mean and unit variance, for example, exponential (1). We then multiply $\zeta_i$ to the weighted data of subject $i$. The resulting perturbed data are $\{(\zeta_i \log(X_{ij}^{(w)}), \zeta_i \boldsymbol{Z}_i^{(w)}); i = 1, \ldots, n, j = 1, \ldots, m_i^*\}$. We then

consider a perturbed objective function for each $\tau_k, k = 1, \ldots, L$,

$$\widetilde{\boldsymbol{V}}_k(\boldsymbol{b}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i^*}\left|\Delta_i\zeta_i\log(X_{ij}^{(w)}) - \Delta_i\zeta_i\boldsymbol{b}^\intercal\boldsymbol{Z}_i^{(w)}\right| + \left|A - \boldsymbol{b}^\intercal\sum_{i=1}^{n}\sum_{j=1}^{m_i^*}(-\Delta_i\zeta_i\boldsymbol{Z}_i^{(w)})\right|$$
$$+ \left|A - \boldsymbol{b}^\intercal\sum_{i=1}^{n}\sum_{j=1}^{m_i^*}\left(2\zeta_i\boldsymbol{Z}_i^{(w)}\sum_{l=0}^{k-1}I\left[\{X_{ij}^{(w)}\}^{\zeta_i} \geqslant \exp\{\zeta_i\boldsymbol{Z}_i^{(w)\intercal}\widetilde{\boldsymbol{\beta}}^*(\tau_l)\}\right] \times \{H(\tau_{l+1}) - H(\tau_l)\}\right)\right|, \quad (7)$$

where $\exp\{\zeta_i\boldsymbol{Z}_i^{(w)\intercal}\widetilde{\boldsymbol{\beta}}^*(0)\}, i = 1, \ldots, n$, are set to be 0. Denote the minimizers of (7) by $\widetilde{\boldsymbol{\beta}}(\tau)$. With the data fixed at the observed values, we repeatedly generate perturbed data a large number of times, say, $B$ times and obtain $B$ realizations, $(\widetilde{\boldsymbol{\beta}}_1^*(\tau), \ldots, \widetilde{\boldsymbol{\beta}}_B^*(\tau))$. Then we can estimate the variance-covariance matrix for $\widehat{\boldsymbol{\beta}}^*(\tau)$ based on the sample statistics of $(\widetilde{\boldsymbol{\beta}}_1^*(\tau), \ldots, \widetilde{\boldsymbol{\beta}}_B^*(\tau))$, based on which the pointwise confidence interval (CI) for $\boldsymbol{\beta}(\tau)$ can be constructed.

To use the bootstrap resampling method, we randomly sample $n$ subjects with replacement from the $n$ subjects in the original data set $B$ times. For each resampled data set, we minimize the target function in (6) with the resampled data to obtain a bootstrap estimate for $\boldsymbol{\beta}(\tau)$. The resulting $B$ realizations of the estimates can then be used to obtain the bootstrap estimate of the variance-covariance matrix for $\widehat{\boldsymbol{\beta}}^*(\tau)$. Note that both resampling methods take subjects as the perturbation or resampling unit, such that the sequence of the gap times of a subject is not disturbed by the resampling procedures.

## 3. Simulations

In this section we investigate the performance of the proposed methods via Monte-Carlo studies, each with 1000 replicates and $n = 100$ within each replicate. We apply the proposed WRS method to the simulated recurrent gap time data, with two different variance estimation methods. For variance estimation, we set the data perturbation times or the bootstrap resampling times $B = 100$. For all the following examples, the censoring times $C_i, i = 1, \ldots, n$

are generated from the uniform distribution on $(0, \pi)$, where $\pi$ is chosen so that the proportion of subjects without any complete gap times is 25% or 40%.

We consider three different settings. In the first example, the regression quantile processes for covariates $Z_1$ is a constant; while in the second example, $\beta_{Z_1}(\tau)$ varies with $\tau$. The third example differs from the second example in the functional form of $\beta_{Z_1}(\tau)$ and an additional covariate $Z_2$.

**Example 1.** We generate recurrent event gap times from an AFT model,

$$\log(T_{ij}) = b_0 + b_1 Z_{1i} + \gamma_i + \epsilon_{ij}, i = 1, \ldots, n, j = 1, 2, \ldots,$$

where $b_0 = -1$, $b_1 = 1$, and $Z_{1i} \sim$ uniform (0, 1). The composite error term $\gamma_i + \epsilon_{ij}$ is composed of the subject-specific random variable $\gamma_i$ and the measurement error $\epsilon_{ij}$, which are independent zero-mean normal variables with variance $\sigma^2 \in (0, 1)$ and $1-\sigma^2$, respectively. Hence, the correlation between two observations $\log(T_{ij})$ and $\log(T_{ij'})$ from the same subject $i$ is $\sigma^2$ if $j \neq j'$. We consider $\sigma^2 = 0.2, 0.4$, and 0.6 for different levels of within-subject correlation between gap times.

The regression quantile for $(1, Z_1)^\intercal$ is given by $\boldsymbol{\beta}(\tau) = (b_0 + Q_{\gamma+\epsilon}(\tau), b_1)^\intercal$, where $Q_{\gamma+\epsilon}(\tau)$, the quantile function of $\gamma_i + \epsilon_{ij}$, is actually the probit function, $\Phi(\tau)$ because $\gamma_i + \epsilon_{ij}$ is a standard normal random variable. Note that in this example the regression quantile process for the covariate $Z$ is a constant, $\beta_{Z_1}(\tau) = b_1$.

**Example 2.** For each subject $i$, $i = 1, \ldots, n$, we generate recurrent event gap times $T_{ij}$ based on a heteroscedastic model with the random errors dependent on the covariate,

$$\log(T_{ij}) = b_0 + b_1 Z_{1i} + (0.5 + Z_{1i})(\gamma_i + \epsilon_{ij}), i = 1, \ldots, n, j = 1, 2, \ldots,$$

where $b_0 = -1$ and $b_1 = 1$. The covariate $Z_{1i}$ and error terms $\gamma_i$ and $\epsilon_{ij}$ are generated the same way as in Example 1. The regression quantile for $(1, Z_1)^\intercal$ is given by $\boldsymbol{\beta}(\tau) = (b_0 + 0.5\Phi(\tau), b_1 + \Phi(\tau))^\intercal$. In this example, the regression quantile process for the covariate $Z_1$ varies with $\tau$, $\beta_{Z_1}(\tau) = b_1 + \Phi(\tau)$. Note that, despite the heteroscedasticity in the error

term $(0.5 + Z_i)(\gamma_i + \epsilon_{ij})$, the correlation between $\log(T_{ij})$ and $\log(T_{ij'})$, $j \neq j'$, is still $\sigma^2$ as in Example 1.

**Example 3.** In this example, $\log(T_{ij})$ follows a multiple-covariate model,

$$\log(T_{ij}) = b_0 + b_1 Z_{1i} + b_2 Z_{2i} + (0.5 + Z_{1i})(\gamma_i + \epsilon_{ij}), i = 1, \ldots, n, j = 1, 2, \ldots,$$

where $b_0 = -1$, $b_1 = 0.5$, and $b_2 = 0.5$. The covariate $Z_{1i} \sim$ uniform (0, 1) and $Z_{2i} \sim$ Bernoulli (0.5). The frailty term $\gamma_i$ is generated the same way as in the previous examples, while $\epsilon_{ij}$ is generated from an independent logistic distribution with zero mean and the scale parameter $s(s > 0)$. The scale parameter satisfies the condition $\sigma^2 + \pi^2 s^2/3 = 1$ so that the composite error term still has a unit variance, and hence the correlation between $\log(T_{ij})$ and $\log(T_{ij'})$, $j \neq j'$, is still $\sigma^2$, as in the previous examples. The regression quantile process is $\boldsymbol{\beta}(\tau) = (b_0 + 0.5Q_{\gamma+\epsilon}(\tau), b_1 + Q_{\gamma+\epsilon}(\tau), b_2)^\intercal$, where $Q_{\gamma+\epsilon}(\tau)$ is the quantile function of the mixture of a normal and a logistic random variable, $\gamma_i + \epsilon_{ij}$ and does not have a closed form expression (Gupta and Nadarajah, 2008).

For comparison, we consider two naive approaches. The first approach is applying the univariate quantile regression model by Peng and Huang (2008) to the first gap time (censored or uncensored) only data (abbreviated as "first gap method"). The second approach is applying the AFT model by Chang (2004) to the recurrent gap time data by assuming a constant regression quantile process for each covariate. Note that, under the model assumption in Example 1, both the quantile regression models and the AFT model are valid methods to use. Tables 1-3 summarize the simulation results for Examples 1-3, respectively. For each scenario, we report the empirical biases (Bias), the empirical standard deviations (SD), the square root of the empirical mean of the estimated variances (SE), and the coverage rate (CR) of the 95% CI for the regression coefficient estimates, for $\tau = 0.2$, 0.4, and 0.6.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

The results in Tables 1-3 show that the point estimates of $\beta_Z(\tau)$ from the proposed methods and the first gap method are approximately unbiased and that the coverage rates of the pointwise 95% CIs are close to the nominal level (0.95). The proposed method is clearly more efficient than the first gap method in all scenarios. In addition, identifiability is found to be less an issue for the proposed method than the first gap method in the sense that the range of $\tau$ satisfying the identifiability condition is wider for the former than the latter in the simulations (not shown). Tables 1-3 also show that the two variance estimation methods, data perturbation and bootstrap resampling methods, perform similarly well for the proposed WRS method. It is also shown that, for a fixed sample size, the smaller the within-subject correlation, the smaller the variance of the estimated regression quantile of the covariate, as would be expected for correlated data. The AFT model estimates are approximately unbiased in Example 1 where the regression quantile process for the covariate is constant. In this case, the AFT model is more efficient than the proposed quantile regression method. For Example 2 with heteroscedastic data, the AFT model gives estimates with large biases. For Example 3 with two covariates, the AFT model estimate is biased for the coefficient of $Z_1$ in the conditional quantile function which changes with $\tau$, while it is approximately unbiased for the coefficient of $Z_2$ which is independent of $Z_1$ and constant with respect to $\tau$.

While it is natural to compare a quantile regression model to an AFT model due to their similar linear form in the conditional quantile function, it has been found less straightforward to compare a quantile regression model with the popular classical PH model, which focuses on the hazard function. The PH model for recurrent gap times by Huang and Chen (2003) takes $\log \lambda$ to be linear in the covariates,

$$\lambda(t|\boldsymbol{Z}_i) = \lambda_0(t) \exp(\boldsymbol{Z}_i^\intercal \boldsymbol{\theta}), \tag{8}$$

where $\boldsymbol{\theta}$ is the regression coefficient, $\lambda_0(t)$ is the baseline hazard function of the gap times

for the subjects whose covariates are zero, and $\lambda(t|\boldsymbol{Z}_i)$ is the hazard function of $T_{ij}$ given $\boldsymbol{Z}_i$, $\lambda(t|\boldsymbol{Z}_i) = -\frac{\mathrm{d}}{\mathrm{d}t}\log S(t|\boldsymbol{Z}_i)$. Following the argument in Portnoy (2003), we can easily prove that under the PH model assumption in (8), the quantile for $T_{ij}$, given $\boldsymbol{Z}_i = \boldsymbol{z}$, is

$$Q_{\mathrm{ph}}(\tau|\boldsymbol{z}) = \Lambda_0^{-1}(-\log(1-\tau)\exp(-\boldsymbol{z}^{\intercal}\boldsymbol{\theta})), \tag{9}$$

where $\Lambda_0(t) \equiv \int_0^t \lambda_0(u)du$ and $\Lambda_0^{-1}(\cdot)$ is the inverse function of $\Lambda_0(\cdot)$. Hence, under the PH model the conditional quantile function is of a very peculiar nonlinear form and it is not easy to make direct comparison. However, Portnoy (2003) and Koenker and Geling (2001) suggested to use a local measure of the effect of the regression coefficient in the PH model on the conditional qantile at $\tau$,

$$b_{\mathrm{ph}}(\tau;\boldsymbol{\theta}) \equiv \frac{\partial}{\partial \boldsymbol{z}} Q_{\mathrm{ph}}(\tau|\boldsymbol{z})\Big|_{\boldsymbol{z}=\bar{\boldsymbol{z}}},$$

which is the derivative of the conditional quantile function in (9) with respect to the covariate $\boldsymbol{z}$, evaluated at the average value of the covariate, $\bar{\boldsymbol{z}}$. The PH regression coefficients, $\boldsymbol{\theta}$ can be estimated by $\widehat{\boldsymbol{\theta}}$ using the method proposed in Huang and Chen (2003). Using the heteroscedastic data generated in Example 2 ($\sigma^2 = 0.4$ and censoring rate = 25%), we compare the proposed regression quantile estimator, $\widehat{\beta}^*(\tau)$ with the analogous PH model effect, $b_{\mathrm{ph}}(\tau;\hat{\boldsymbol{\theta}})$. In Figure 2, the proposed estimator (left panel) and the PH model effect (right panel) are presented as dashed lines and compared with the true regression quantile coefficient $b_1+\Phi(\tau)$ in solid lines. Clearly, the proposed estimator is virtually unbiased, whereas the PH model estimate differs significantly from the true regression quantile coefficient for a wide range of $\tau$ and has a wider confidence interval in the upper tail of $\tau$.

[Figure 2 about here.]

## 4. An Example

To illustrate the proposed methods, we analyze a subset of the Danish Psychiatric Central Register data (Munk-Jørgensen and Mortensen, 1997), described in Luo and Huang (2011).

The dataset consists of 286 individuals who were first admitted to, or had contacts with Danish psychiatric services during the period from April 1, 1970 to December 31, 1970 with a diagnosis of schizophrenia. We set the maximum follow-up time for each person to be 3 years to avoid the potential change in the distributional pattern of recurrent gap times. Among the 286 individuals, 106 (37%) were females, 115 (40%) did not have another hospitalization after the first hospitalization or contact, 47 (20%) had only one re-hospitalization and the rest 115 (40%) had more than one re-hospitalization. On average, each patient experienced 1.7 hospitalizations after the initial hospitalization or contact. Nine patients in this cohort died before the end of follow-up. Hence, the assumption of independent censoring is not expected to be seriously violated.

We are interested in assessing the effect of the onset age of schizophrenia on the length between two adjacent hospitalizations due to a schizophrenic episode. The onset age of this cohort ranges from 14 to 88, and the three quartiles are (21, 26, 39). A two-sample test comparing those who had schizophrenia onset as early as or before 20 years old (early onset) versus after 20 years old (late onset) and a proportional hazards regression analysis for the continuous onset age were reported in Luo and Huang (2011). Both analyses revealed that late onset ages of schizophrenia were significantly associated with longer gap times between hospitalizations. We apply the proposed WRS method to fit quantile regression models to evaluate the effect of age of onset (in years) on the quantiles of gap times, adjusting for gender. Figure 3 depicts the estimated regression quantile process for age of onset (left panel) and gender (right panel) based on different methods. Their associated 95% pointwise CI's are presented in Table 4. The estimated quantile process for onset age based on the WRS method is positive, ranging between 0.03 and 0.09, for $\tau$ between 0.1 and 0.6, suggesting that later onset ages are associated with longer gap times between two adjacent hospitalizations. We observe that the estimated quantile is virtually constant when $\tau$ ranges between 0.1 and

0.45, but after 0.45 it starts to increase steadily. The estimated regression quantile coefficient based on the first gap method (shown as dashed lines in Figure 3) indicates a similar effect of onset age as the WRS method, whereas the 95% pairwise confidence interval is slightly wider than that of the WRS method. The estimate of the onset age effect based on the AFT model (Chang, 2004), depicted as the dot-dashed line in Figure 3, falls within the range of the maximum and minimum values of the estimated regression quantile coefficients of the WRS method or the first gap method, but it clearly differs from the estimated regression quantile coefficient for $\tau$ close to or larger than 0.5. The effect of gender, in the presence of onset age in the regression, is not significant in any of the models.

[Figure 3 about here.]

[Table 4 about here.]

## 5. Remarks

Quantile regression has received much attention recently. Most existing censored quantile regression methods focus on univariate survival data. The limited work on multivariate survival data includes Yin and Cai (2005) which studied quantile regression for clustered survival data. In this article, we study quantile regression models for recurrent gap time data, which can serve as a useful alternative to hazard-based models and mean-based models. To estimate the regression quantile process for recurrent gap time data with the proposed method, the function *crq* in R package *quantreg* can be used directly on a pre-cleaned working data set where the last censored gap should be removed for those at least one complete gap time has been observed. A weight vector $((m_1^*)^{-1}\mathbf{1}^\intercal_{m_1^*\times 1},\ldots,(m_n^*)^{-1}\mathbf{1}^\intercal_{m_n^*\times 1})^\intercal$, where $\mathbf{1}$ is a vector of one, should be provided to the R function *crq*. Modifications on the R functions *summary.crq* and *boot.crq* in the same R package are needed so that the perturbation or resmapling schemes are performed on subjects rather than on individual gap times. The

sequence of the gap times from the same subject, hence, can be preserved when estimating the variance or confidence interval of the regression quantile process.

The proposed WRS estimator for the regression quantile coefficients can be approximated by applying the modified within-cluster resampling (MWCR) scheme (Luo and Huang, 2011) on the univariate quantile regression by Peng and Huang (2008). Specifically, for each resampling, one gap time is randomly selected from the first $m_i^*$ gap times for each person. Let $J(i) \in \{1, 2, \ldots, m_i^*\}$ be the index for the selected gap time for subject $i$. The data from the $b$th resampling, $b = 1, \ldots, M$, are i.i.d. and denoted by $\{(X_{iJ_b(i)}, \Delta_i, \boldsymbol{Z}_i), i = 1, \ldots, n\}$. Let $\widetilde{\boldsymbol{\beta}}_b(\tau)$ and $\widetilde{\boldsymbol{\Sigma}}_b(\tau), \tau \in (0, 1)$ be the estimated quantile regression process and the corresponding variance-covariance process based on the $b$th resampled data by using Peng and Huang's method. The MWCR estimator for $\boldsymbol{\beta}(\tau)$ is $\widetilde{\boldsymbol{\beta}}(\tau) = M^{-1} \sum_{b=1}^{M} \widetilde{\boldsymbol{\beta}}_b(\tau)$ and the variance of $\widetilde{\boldsymbol{\beta}}(\tau)$ can be estimated by $\widetilde{\boldsymbol{\Sigma}}(\tau) = \frac{1}{M} \sum_{b=1}^{M} \widetilde{\boldsymbol{\Sigma}}_b(\tau) - \frac{1}{M-1} \sum_{b=1}^{M} \left( \widetilde{\boldsymbol{\beta}}_b(\tau) - \widetilde{\boldsymbol{\beta}}(\tau) \right)^{\otimes 2}$.

As suggested by a referee and also noted by Wang and Chang (1999), the efficiency gain of the WRS-based methods relative to the first gap method may be diminished if the correlation among recurrent gap times is strong. More efficient estimators can be constructed by incorporating weight functions into estimating equations, that is,

$$\boldsymbol{U}^a(\boldsymbol{\beta}, \tau) = n^{-1} \sum_{i=1}^{n} a_i \boldsymbol{Z}_i \left[ N_i^* \left( \exp\{\boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(\tau)\} \right) - \int_0^\tau R_i^* \left( \exp\{\boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(u)\} \right) dH(u) \right], \qquad (10)$$

where $a_i$ is a positive-valued and bounded weight function subject to the constraint $\mathrm{E}(a_i^2) < \infty$. Alternatively, $\boldsymbol{U}^a(\boldsymbol{\beta}, \tau)$ can be expressed as

$$n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i \left[ N_i^a \left( \exp\{\boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(\tau)\} \right) - \int_0^\tau R_i^a \left( \exp\{\boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(u)\} \right) dH(u) \right],$$

where $N_i^a(t) = (a_i / m_i^*) \sum_{j=1}^{m_i^*} N_{ij}(t)$ and $R_i^a(t) = (a_i / m_i^*) \sum_{j=1}^{m_i^*} R_{ij}(t)$. Similarly, $N_i^a(t)$ and $R_i^a(t)$ can be used in other WRS-based methods discussed in Luo and Huang (2011) to improve efficiency. By simple algebra, we can prove that $\boldsymbol{U}^a(\boldsymbol{\beta}, \tau)$ is an unbiased estimating function, which ensures the consistency of the estimator from solving $\boldsymbol{U}^a(\boldsymbol{\beta}, \tau) = 0$. Obviously, a convenient choice of $a_i = 1$ results in the estimating function $\boldsymbol{U}^*(\boldsymbol{\beta}, \tau)$. Theoretically,

the optimal weighted average can be achieved by using the weights which are proportional to the inverse of the variances of the summand, $\boldsymbol{Z}_i \left[ N_i^* \left( \exp\{ \boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(\tau) \} \right) - \int_0^\tau R_i^* \left( \exp\{ \boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(u) \} \right) dH(u) \right]$ in (10). To this end, techniques in estimating the within-subject correlation structure among the gap times are needed and will be exploited in our future research.

The proposed model in (1) assumes that the linearity between the covariates and the conditional quantiles hold for all $\tau \in (0, 1)$. This global linearity assumption may not hold in practice. For univariate survival data, Peng and Huang (2008) considered a class of supremum test statistics based on the large sample distribution of a functional of the martingale residual processes, $M_{i1}(\tau; \boldsymbol{\beta}) = \left[ N_{i1} \left( \exp\{ \boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(\tau) \} \right) - \int_0^\tau R_{i1} \left( \exp\{ \boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(u) \} \right) dH(u) \right]$. Along the lines of Huang, Luo, and Follmann (2011), we can construct similar test statistics for the quantile regression model in (1) for recurrent gap time by replacing $M_{i1}(\tau; \boldsymbol{\beta})$ with $M_i^*(\tau; \boldsymbol{\beta}) = \left[ N_i^* \left( \exp\{ \boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(\tau) \} \right) - \int_0^\tau R_i^* \left( \exp\{ \boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(u) \} \right) dH(u) \right]$ and construct test statistics based on $M_i^*(\tau; \boldsymbol{\beta})$. In the case when the global linearity assumption in (1) is violated, a naive approach would be applying the locally weighted quantile regression by Wang and Wang (2009) on the first gap times. Further research in extending quantile regression for recurrent gap time which only requires local linearity is certainly warranted.

The validity of the proposed estimation procedure relies on the exchangeability property implied by the conditionally i.i.d. assumption (M1) imposed in this paper and many others, including Huang and Chen (2003), Chang (2004), and Lu (2005). The conditional frailty model is useful in describing heterogeneity due to omitted covariates – thus gap times from the same subject are more similar than those from different subjects. However, we acknowledge that the exchangeable correlation structure can be inadequate when the gap times within a subject is complex. The readers are referred to Cook and Lawless (2006) for discussions on the limitations of the exchangeable correlation structure, as well as for

an alternative parametric log-normal model for recurrent gap times to allow for flexible correlation structures.

The method discussed in this paper is only applicable for studying the effects of covariates when different episodes of uncensored recurrent gap times within a subject have the same distribution. Hence, when this condition fails to hold, the proposed method may yield biased inferential results. We suggest that before using the proposed quantile regression, one can first use the nonparametric trend test proposed by Wang and Chen (2000) to test if there exists any change in the distributional pattern of the recurrent gap times. It will be interesting for future research to exploit models that allow episode-specific regression quantile coefficients (i.e., $\boldsymbol{\beta}_j(\tau)$) or covariates (i.e., $\boldsymbol{Z}_{ij}$).

## Acknowledgements

We would like to thank the Associate Editor and reviewer for their valuable comments, Drs. Preben Bo Mortensen and William Eaton for providing schizophrenia data, Mr. Todd DeFor for editing manuscript drafts, and University of Minnesota Supercomputing Institute for providing computing resources for this research. This research was supported by University of Minnesota Grant-in-Aid #22281 to Luo, by grant NSF DMS-1007603 to Wang.

## Supplementary Materials

Web Appendices A and B, referenced in Section 2.2, the data example, used in 4, and the R code, referenced in Section 5 are available with this paper at the Biometrics website on Wiley Online Library.

## References

Alexander, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Annals of Probability* **12,** 1041–1067.

Chang, S.-H. (2004). Estimating marginal effects in accelerated failure time models for serial sojourn times among repeated events. *Lifetime Data Analysis* **10,** 175–190.

Cook, R. J. and Lawless, J. F. *The Statistical Analysis of Recurrent Events.* Springer: New York, NY, 2007.

Fygenson, M., and Ritov, Y. (1994). Monotone estimating equations for censored data. *The Annals of Statistics* **22,** 732–746.

Gupta, A. K. and Nadarajah, S. (2008). Normal and logistic random variables: distribution of the linear combination. *Statistical Papers* **49,** 201–209.

Huang, C.-Y., Luo, X., and Follmann, D. A. (2011). A model checking method for the proportional hazards model with recurrent gap time data. *Biostatistics* **12,** 535-547.

Huang, X. and Liu, L. A joint frailty model for survival and gap times between recurrent events. (2007) **Biometrics** *63,* 389–397.

Huang, Y. J. and Chen, Y. Q. (2003). Marginal regression of gaps between recurrent events. *Lifetime Data Analysis* **9,** 293–303.

Jin, Z., Ying, Z., and Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88,** 381–390.

Koenker, R. (2008a). Censored quantile regression redux. *Journal of Statistical Software* **27,** 1-25.

Koenker, R. (2008b). *quantreg*: quantile regression, R package version 4.16, available at http://www.r-project.org.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46,** 33–50.

Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association* **96,** 458–468.

Lai, T. L. and Ying Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regression. *Journal of Multivariate Analysis* **27,** 334–358.

Lu, W. (2005). Marginal regression of multivariate event times based on linear transformation models. *Lifetime Data Analysis* **11,** 389–404.

Luo, X. and Huang, C.-Y. (2011). Analysis of recurrent gap time data using the weighted risk set method and the modified within-cluster resampling method. *Statistics in Medicine* **30,** 301–311.

Munk-Jørgensen, P. and Mortensen, P. B. (1997). The Danish Psychiatric Central Register. *Danish Medical Bulletin* **44,** 82–84.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of American Statistical Association* **103,** 637–649.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association* **98,** 1001–1012.

Powell, J. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* **25,** 303–325.

Powell, J. (1986). Censored regression quantiles. *Journal of Econometrics* **32,** 143–155.

Strawderman, R. L. (2005). The accelerated gap times model. *Biometrika* **92,** 647–666.

Sun, L. Q., Park, D. H., and Sun, J. G. (2006). The additive hazards model for recurrent gap times. *Statistica Sinica* **16,** 919–932.

Wang, H. J. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association* **104,** 1117–1128.

Wang, M.-C. and Chang, S.-H. (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association* **94,** 146–153.

Wang, M.-C. and Chen, Y. Q. (2000). Nonparametric and semiparametric trend analysis for stratified recurrence time data. *Biometrics* **56,** 789–794.

Yang, S. (1999). Censoring median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association* **94,** 137–145.

Yin, G. and Cai, J. (2005). Quantile regression models with multivariate failure time data. *Biometrics* **61,** 151–161.

Ying, Z., Jung, S. H., and Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90,** 178–184.

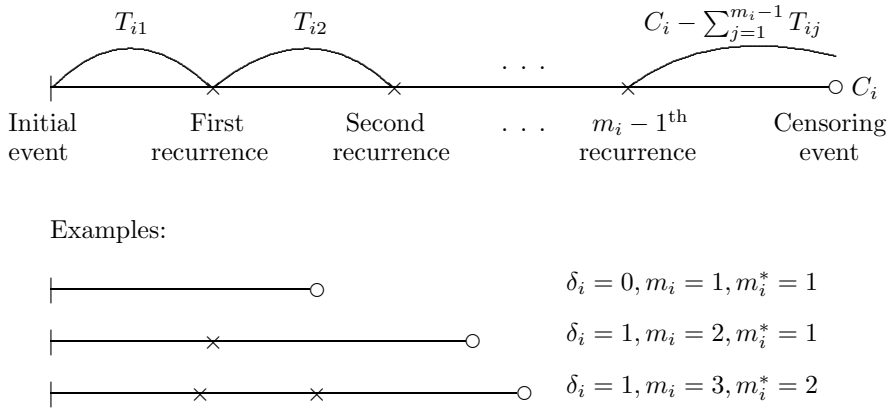


**Figure 1.** Illustration of recurrent gap time data

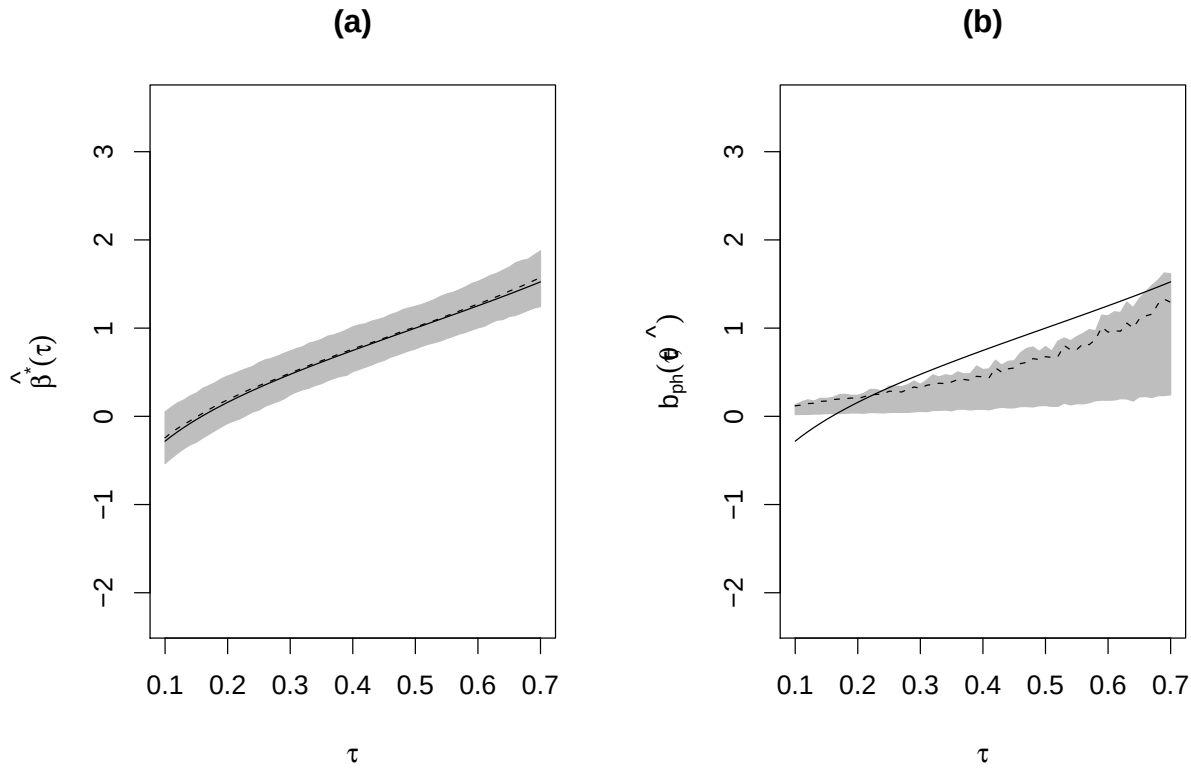


**Figure 2.** Comparison of the proposed quantile regression coefficient estimator (panel a) and the proportional hazards (PH) model effect (panel b) for simulated data. The solid line in both panels is the true regression quantile coefficient for $Z_1$, $b_1 + \Phi(\tau)$; the dashed lines are for the proposed estimator (in panel a) and the PH model effect (in panel b), respectively; and the shaded areas are their corresponding empirical 95% pointwise confidence intervals.


**Figure 3.** The estimated regression quantile for onset age of schizophrenia and gender for $0.1 \leqslant \tau \leqslant 0.6$. The solid line is the estimated regression quantile based on the weighted risk-set method; the dashed line is the first gap method estimate; and the dot-dashed line is the accelerated failure time model estimate.

**Table 1**
*Summary of the simulation results for Example 1*

| | | | First gap method | | | | AFT model | | | | Proposed method | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | Perturb | | Bootstrap | |
| $\sigma^2$ | $\bar{m}$ [a] | $\tau$ | Bias [b] | SD [c] | SE [d] | CR [e] | Bias | SD | SE | CR | Bias | SD | SE | CR | SE | CR |
| Censoring rate[f] = 25% | | | | | | | | | | | | | | | | |
| 0.2 | 3.42 | 0.2 | .006 | .520 | .525 | .925 | .011 | .362 | .366 | .925 | .022 | .363 | .397 | .932 | .386 | .937 |
| | | 0.4 | -.004 | .467 | .481 | .942 | .011 | .362 | .366 | .925 | .004 | .371 | .379 | .936 | .381 | .935 |
| | | 0.6 | .004 | .479 | .512 | .937 | .011 | .362 | .366 | .925 | .007 | .406 | .431 | .933 | .426 | .937 |
| 0.4 | 3.78 | 0.2 | .005 | .516 | .534 | .926 | .025 | .378 | .377 | .925 | .012 | .380 | .383 | .928 | .392 | .934 |
| | | 0.4 | .002 | .463 | .478 | .937 | .025 | .378 | .377 | .925 | .014 | .372 | .385 | .944 | .390 | .947 |
| | | 0.6 | .027 | .477 | .513 | .944 | .025 | .378 | .377 | .925 | .014 | .418 | .435 | .946 | .441 | .942 |
| 0.6 | 4.28 | 0.2 | .001 | .501 | .528 | .937 | .018 | .394 | .384 | .923 | .015 | .403 | .396 | .923 | .402 | .931 |
| | | 0.4 | -.006 | .465 | .470 | .933 | .018 | .394 | .384 | .923 | .010 | .395 | .392 | .936 | .394 | .934 |
| | | 0.6 | .007 | .495 | .508 | .934 | .018 | .394 | .384 | .923 | .013 | .442 | .444 | .930 | .447 | .941 |
| Censoring rate = 40% | | | | | | | | | | | | | | | | |
| 0.2 | 2.27 | 0.2 | .039 | .506 | .553 | .919 | .012 | .411 | .406 | .931 | .026 | .434 | .462 | .913 | .466 | .918 |
| | | 0.4 | .026 | .488 | .533 | .918 | .012 | .411 | .406 | .931 | .023 | .448 | .473 | .917 | .478 | .932 |
| | | 0.6 | .060 | .721 | .680 | .923 | .012 | .411 | .406 | .931 | .039 | .592 | .601 | .921 | .624 | .932 |
| 0.4 | 2.44 | 0.2 | .038 | .518 | .560 | .914 | .012 | .402 | .411 | .929 | .023 | .437 | .460 | .913 | .468 | .928 |
| | | 0.4 | .010 | .485 | .525 | .922 | .012 | .402 | .411 | .929 | .009 | .438 | .463 | .921 | .472 | .927 |
| | | 0.6 | .025 | .566 | .649 | .919 | .012 | .402 | .411 | .929 | .024 | .535 | .580 | .913 | .593 | .920 |
| 0.6 | 2.69 | 0.2 | .023 | .493 | .565 | .925 | .017 | .407 | .418 | .922 | .014 | .434 | .467 | .909 | .478 | .917 |
| | | 0.4 | .014 | .482 | .525 | .927 | .017 | .407 | .418 | .922 | .022 | .447 | .475 | .919 | .481 | .921 |
| | | 0.6 | .053 | .567 | .638 | .939 | .017 | .407 | .418 | .922 | .047 | .552 | .620 | .931 | .614 | .944 |

[a] Average number of censored and uncensored gap times per subject
[b] Empirical bias
[c] Empirical standard deviation
[d] Square root of the empirical mean of the estimated variances
[e] Coverage rate of the nominal 95% confidence intervals
[f] Proportion of subjects without any complete gap times

**Table 2**
*Summary of the simulation results for Example 2*

| | | | First gap method | | | | AFT model | | | | Proposed method | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | Perturb | | Bootstrap | |
| $\sigma^2$ | $\bar{m}$ [a] | $\tau$ | Bias [b] | SD [c] | SE [d] | CR [e] | Bias | SD | SE | CR | Bias | SD | SE | CR | SE | CR |
| Censoring rate[f] = 25% | | | | | | | | | | | | | | | | |
| 0.2 | 3.52 | 0.2 | -.003 | .408 | .411 | .937 | 1.161 | .360 | .351 | .095 | .022 | .317 | .309 | .921 | .311 | .933 |
| | | 0.4 | .024 | .369 | .378 | .933 | .573 | .360 | .351 | .592 | .026 | .311 | .306 | .925 | .308 | .931 |
| | | 0.6 | .033 | .394 | .400 | .926 | .066 | .360 | .351 | .916 | .024 | .338 | .340 | .933 | .343 | .934 |
| 0.4 | 3.82 | 0.2 | -.006 | .410 | .406 | .920 | 1.147 | .349 | .360 | .112 | .017 | .314 | .306 | .923 | .310 | .926 |
| | | 0.4 | .019 | .373 | .374 | .927 | .559 | .349 | .360 | .640 | .012 | .306 | .308 | .941 | .310 | .936 |
| | | 0.6 | .024 | .374 | .402 | .937 | .052 | .349 | .360 | .929 | .017 | .329 | .352 | .943 | .355 | .945 |
| 0.6 | 4.29 | 0.2 | -.010 | .394 | .414 | .935 | 1.155 | .352 | .362 | .120 | .015 | .313 | .314 | .925 | .320 | .929 |
| | | 0.4 | -.001 | .357 | .371 | .939 | .567 | .352 | .362 | .616 | .006 | .302 | .313 | .944 | .317 | .937 |
| | | 0.6 | .011 | .375 | .397 | .939 | .060 | .352 | .362 | .926 | .014 | .345 | .354 | .941 | .354 | .937 |
| Censoring rate = 40% | | | | | | | | | | | | | | | | |
| 0.2 | 2.24 | 0.2 | .041 | .490 | .548 | .916 | .942 | .364 | .363 | .291 | .044 | .424 | 460 | .911 | .469 | .912 |
| | | 0.4 | .030 | .487 | .521 | .925 | .354 | .364 | .363 | .801 | .056 | .461 | .477 | .905 | .486 | .919 |
| | | 0.6 | .086 | .602 | .663 | .934 | -.153 | .364 | .363 | .902 | .081 | .569 | .595 | .907 | .623 | .928 |
| 0.4 | 2.40 | 0.2 | .036 | .508 | .550 | .925 | .933 | .363 | .369 | .297 | .049 | .441 | .453 | .920 | .467 | .921 |
| | | 0.4 | .018 | .476 | .519 | .918 | .344 | .363 | .369 | .813 | .045 | .440 | .470 | .911 | .478 | .916 |
| | | 0.6 | .071 | .576 | .675 | .919 | -.162 | .363 | .369 | .897 | .073 | .546 | .623 | .910 | .628 | .927 |
| 0.6 | 2.61 | 0.2 | .037 | .501 | .549 | .921 | .940 | .370 | .374 | .284 | .047 | .439 | .462 | .920 | .472 | .920 |
| | | 0.4 | .023 | .479 | .519 | .915 | .351 | .370 | .374 | .802 | .045 | .445 | .481 | .919 | .484 | .921 |
| | | 0.6 | .103 | .614 | .660 | .922 | -.155 | .370 | .374 | .914 | .095 | .565 | .620 | .928 | .640 | .929 |

[a] Average number of censored and uncensored gap times per subject
[b] Empirical bias
[c] Empirical standard deviation
[d] Square root of the empirical mean of the estimated variances
[e] Coverage rate of the nominal 95% confidence intervals
[f] Proportion of subjects without any complete gap times

**Table 3**
*Summary of the simulation results for Example 3*

| | | | | First gap method | | | | AFT model | | | | Proposed method | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | Perturb | | Bootstrap | |
| $\sigma^2$ | $\bar{m}$ [a] | | $\tau$ | Bias [b] | SD [c] | SE [d] | CR [e] | Bias | SD | SE | CR | Bias | SD | SE | CR | SE | CR |
| | | | | | | | | $\beta_1(\tau) = b_1 + Q_{\gamma+\epsilon}(\tau)$ | | | | | | | | | |
| Censoring rate[f] = 25% | | | | | | | | | | | | | | | | | |
| 0.2 | 3.44 | | 0.2 | .031 | .488 | .520 | .913 | 1.111 | .349 | .352 | .116 | .049 | .366 | .382 | .927 | .397 | .939 |
| | | | 0.4 | .028 | .409 | .465 | .939 | .566 | .349 | .352 | .648 | .046 | .331 | .367 | .931 | .378 | .942 |
| | | | 0.6 | .042 | .425 | .484 | .940 | .080 | .349 | .352 | .931 | .051 | .380 | .408 | .932 | .418 | .941 |
| 0.4 | 3.81 | | 0.2 | .017 | .509 | .534 | .919 | 1.141 | .358 | .366 | .130 | .037 | .391 | .398 | .912 | .410 | .920 |
| | | | 0.4 | .027 | .438 | .474 | .922 | .578 | .358 | .366 | .612 | .040 | .374 | .383 | .921 | .395 | .925 |
| | | | 0.6 | .010 | .457 | .493 | .927 | .083 | .358 | .366 | .939 | .035 | .401 | .426 | .920 | .439 | .944 |
| 0.6 | 4.33 | | 0.2 | .052 | .498 | .527 | .911 | 1.177 | .376 | .370 | .122 | .059 | .398 | .408 | .914 | .419 | .926 |
| | | | 0.4 | .061 | .450 | .488 | .917 | .605 | .376 | .370 | .607 | .069 | .385 | .401 | .931 | .411 | .928 |
| | | | 0.6 | .051 | .458 | .502 | .927 | .107 | .376 | .370 | .916 | .068 | .414 | .440 | .924 | .448 | .926 |
| Censoring rate = 40% | | | | | | | | | | | | | | | | | |
| 0.2 | 2.26 | | 0.2 | .036 | .521 | .546 | .922 | .926 | .370 | .369 | .322 | .046 | .439 | .459 | .920 | .470 | .915 |
| | | | 0.4 | .045 | .444 | .492 | .926 | .381 | .370 | .369 | .798 | .054 | .408 | .440 | .938 | .447 | .931 |
| | | | 0.6 | .057 | .518 | .570 | .921 | -.104 | .370 | .369 | .897 | .062 | .492 | .524 | .915 | .543 | .925 |
| 0.4 | 2.43 | | 0.2 | .025 | .515 | .549 | .916 | .932 | .368 | .379 | .330 | .032 | .430 | .456 | .915 | .474 | .924 |
| | | | 0.4 | .016 | .470 | .505 | .925 | .370 | .368 | .379 | .815 | .039 | .434 | .446 | .914 | .456 | .933 |
| | | | 0.6 | .041 | .544 | .587 | .926 | -.125 | .368 | .379 | .912 | .055 | .520 | .550 | .931 | .561 | .941 |
| 0.6 | 2.68 | | 0.2 | .046 | .513 | .550 | .917 | .098 | .371 | .386 | .297 | .063 | .446 | .454 | .915 | .469 | .921 |
| | | | 0.4 | .045 | .465 | .516 | .937 | .405 | .371 | .386 | .795 | .063 | .429 | .463 | .924 | .475 | .928 |
| | | | 0.6 | .084 | .547 | .595 | .936 | -.092 | .371 | .386 | .918 | .088 | .528 | .556 | .926 | .575 | .940 |
| | | | | | | | | $\beta_2(\tau) = b_2$ | | | | | | | | | |
| Censoring rate = 25% | | | | | | | | | | | | | | | | | |
| 0.2 | 3.44 | | 0.2 | .006 | .266 | .287 | .922 | -.022 | .200 | .198 | .929 | .002 | .191 | .206 | .948 | .211 | .945 |
| | | | 0.4 | .000 | .236 | .250 | .944 | -.022 | .200 | .198 | .929 | -.005 | .185 | .194 | .939 | .197 | .940 |
| | | | 0.6 | -.010 | .242 | .261 | .940 | -.022 | .200 | .198 | .929 | -.003 | .200 | .214 | .936 | .216 | .943 |
| 0.4 | 3.81 | | 0.2 | -.004 | .272 | .290 | .932 | -.014 | .202 | .204 | .939 | .007 | .202 | .211 | .945 | .219 | .947 |
| | | | 0.4 | -.006 | .240 | .256 | .937 | -.014 | .202 | .204 | .939 | -.001 | .196 | .204 | .951 | .207 | .942 |
| | | | 0.6 | -.005 | .247 | .263 | .931 | -.014 | .202 | .204 | .939 | .000 | .214 | .226 | .937 | .228 | .944 |
| 0.6 | 4.33 | | 0.2 | -.014 | .272 | .291 | .937 | -.028 | .209 | .208 | .920 | -.012 | .211 | .219 | .937 | .225 | .942 |
| | | | 0.4 | -.009 | .248 | .262 | .925 | -.028 | .209 | .208 | .920 | -.009 | .208 | .214 | .935 | .216 | .938 |
| | | | 0.6 | -.002 | .253 | .267 | .944 | -.028 | .209 | .208 | .920 | -.007 | .226 | .231 | .935 | .235 | .938 |
| Censoring rate = 40% | | | | | | | | | | | | | | | | | |
| 0.2 | 2.26 | | 0.2 | -.014 | .271 | .293 | .934 | -.036 | .216 | .209 | .914 | -.006 | .229 | .245 | .943 | .252 | .955 |
| | | | 0.4 | -.008 | .247 | .267 | .936 | -.036 | .216 | .209 | .914 | -.006 | .222 | .231 | .936 | .237 | .934 |
| | | | 0.6 | .002 | .269 | .280 | .918 | -.036 | .216 | .209 | .914 | -.001 | .256 | .257 | .922 | .262 | .923 |
| 0.4 | 2.43 | | 0.2 | -.007 | .274 | .299 | .935 | -.020 | .206 | .215 | .936 | .007 | .231 | .245 | .943 | .250 | .947 |
| | | | 0.4 | -.001 | .250 | .274 | .952 | -.020 | .206 | .215 | .936 | .003 | .227 | .242 | .938 | .246 | .950 |
| | | | 0.6 | .005 | .258 | .284 | .923 | -.020 | .206 | .215 | .936 | .005 | .253 | .265 | .922 | .273 | .940 |
| 0.6 | 2.68 | | 0.2 | -.004 | .273 | .301 | .935 | -.027 | .212 | .220 | .932 | -.001 | .235 | .247 | .934 | .253 | .939 |
| | | | 0.4 | .002 | .248 | .273 | .941 | -.027 | .212 | .220 | .932 | .004 | .230 | .247 | .941 | .251 | .945 |
| | | | 0.6 | .002 | .268 | .296 | .938 | -.027 | .212 | .220 | .932 | .000 | .258 | .277 | .945 | .280 | .943 |

[a] Average number of censored and uncensored gap times per subject
[b] Empirical bias
[c] Empirical standard deviation
[d] Square root of the empirical mean of the estimated variances
[e] Coverage rate of the nominal 95% confidence intervals
[f] Proportion of subjects without any complete gap times

**Table 4**
*Summary of estimated regression quantile for schizophrenia data.*

| | First gap method | | | AFT model | | | Proposed method | | |
|---|---|---|---|---|---|---|---|---|---|
| $\tau$ | $\beta(\tau)$[a] | s.e. [b] | 95% CI [c] | $\beta(\tau)$ | s.e. | 95% CI | $\beta(\tau)$ | s.e. | 95% CI |
| | | | | Age of onset | | | | | |
| 0.1 | .040 | .011 | (.019, .062) | .042 | .008 | (.027, .057) | .035 | .010 | (.016, .053) |
| 0.2 | .033 | .012 | (.009, .057) | .042 | .008 | (.027, .057) | .030 | .008 | (.015, .045) |
| 0.3 | .032 | .013 | (.007, .060) | .042 | .008 | (.027, .057) | .032 | .012 | (.009, .055) |
| 0.4 | .042 | .009 | (.023, .060) | .042 | .008 | (.027, .057) | .036 | .009 | (.018, .053) |
| 0.5 | .105 | .044 | (.019, .191) | .042 | .008 | (.027, .057) | .075 | .036 | (.005, .145) |
| 0.6 | .117 | .038 | (.042, .192) | .042 | .008 | (.027, .057) | .093 | .030 | (.034, .153) |
| | | | | Gender (female) | | | | | |
| 0.1 | -.211 | .539 | (-1.267, .845) | .070 | .234 | (-.455, .469) | -.059 | .346 | (-.737, .619) |
| 0.2 | -.163 | .360 | (-.868, .541) | .070 | .234 | (-.455, .469) | .054 | .349 | (-.630, .739) |
| 0.3 | .116 | .288 | (-.449, .681) | .070 | .234 | (-.455, .469) | .192 | .251 | (-.300, .683) |
| 0.4 | -.042 | .255 | (-.541, .457) | .070 | .234 | (-.455, .469) | .148 | .175 | (-.195, .490) |
| 0.5 | .150 | .413 | (-.660, .960) | .070 | .234 | (-.455, .469) | .128 | .394 | (-.645, .900) |
| 0.6 | .173 | .352 | (-.517, .864) | .070 | .234 | (-.455, .469) | .332 | .237 | (-.133, .797) |

[a] Point estimate based on different methods
[b] Standard error (the data perturbation method was used for the first gap method and the proposed method)
[c] 95% pointwise confidence interval